Molecular Dynamics and Free Energy Perturbation Study of Hydride-Ion Transfer Step in Dihydrofolate Reductase Using Combined Quantum and Molecular Mechanical Model

PETER L. CUMMINS, JILL E. GREADY

Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, ACT 2601, Australia

Received 24 October 1997; accepted 28 January 1998

ABSTRACT: We used molecular dynamics simulation and free energy perturbation (FEP) methods to investigate the hydride-ion transfer step in the mechanism for the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of a novel substrate by the enzyme dihydrofolate reductase (DHFR). The system is represented by a coupled quantum mechanical and molecular mechanical (QM/MM) model based on the AM1 semiempirical molecular orbital method for the reacting substrate and NADPH cofactor fragments, the AMBER force field for DHFR, and the TIP3P model for solvent water. The FEP calculations were performed for a number of choices for the QM system. The substrate, 8-methylpterin, was treated quantum mechanically in all the calculations, while the larger cofactor molecule was partitioned into various QM and MM regions with the addition of ''link'' atoms (F, CH_3, and H). Calculations were also carried out with the entire NADPH molecule treated by QM. The free energies of reaction and the net charges on the NADPH fragments were used to determine the most appropriate QM/MM model. The hydride-ion transfer was also carried out over several FEP pathways, and the QM and QM/MM component free energies thus calculated were found to be state functions (i.e., independent of pathway). A ca. 10 kcal/mol increase in free energy for the hydride-ion transfer with an activation barrier of ca. 30 kcal/mol was calculated. The increase in free energy on the hydride-ion transfer arose largely from the QM/MM component. Analysis of the QM/MM energy components suggests that, although a number of charged residues may contribute to the free energy change through long-range electrostatic interactions,

Correspondence to: P. L. Cummins
Contract/grant sponsor: Australian National University

© 1998 John Wiley & Sons, Inc.
CCC 0192-8651 / 98 / 080977-12

the only interaction that can account for the 10 kcal/mol increase in free energy is the hydrogen bond between the carboxylate side chain of Glu30 (avian DHFR) and the activated (protonated) substrate. © 1998 John Wiley & Sons, Inc.
J Comput Chem 19: 977–988, 1998

Keywords: quantum mechanical/molecular mechanical; free energy; hydride ion; molecular dynamics; catalysis

Introduction

To be effective catalysts, enzymes must somehow reduce the activation free energy of the corresponding chemical reaction that occurs in solution. The key question is how enzymes achieve this. It is difficult to define enzyme catalysis at the atomic level using experimental methods; consequently, many competing hypotheses have been advanced to explain enzymic mechanisms and energetics. Increasingly, computational methods have been employed to study the mechanisms in an effort to test the validity of these ideas.[1–5] Of particular interest are theoretical studies that have implicated specific electrostatic interactions between the transition state (TS) and the protein as the main factors accounting for enzyme catalysis.[1,3] In effect, this implies that nature has engineered the disposition of the polar and charged groups in the enzyme active site toward TS stabilization for a particular reaction. However, the extent to which such electrostatic-stabilization mechanisms apply is not known, because there are clearly other mechanisms that could explain the activation of a chemical reaction.

Due largely to its importance as a target for anticancer and antibacterial drugs, dihydrofolate reductase (DHFR) is an enzyme that continues to attract considerable interest, particularly with respect to elucidation of the catalytic mechanism.[6] Indeed, computational methods have been an important component in the study of this mechanism.[6–11] DHFR catalyzes the hydride-ion transfer between the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor and a substrate molecule (S) according to

$$\mathrm{NADPH + S + H^+ \rightarrow NADP^+ + SH_2},$$

where the normal substrates are folate and dihydrofolate. Note that the hydride-ion transfer is pH dependent as it requires the transfer of a proton to the substrate. Because the natural fully oxidized substrate folate is unprotonated when bound to the enzyme at physiological pH, how the proton finds its way to this substrate in the active site appears to be critical to an understanding of the mechanism of activation toward hydride-ion transfer.[6] However, we developed a series of novel substrates, the 8-substituted pterins, that, unlike folate, have a high pK_a (ca. 5.5) in solution and have been predicted[12] to bind strongly to the active site in the protonated form through charge–charge interactions. In fact, the results obtained from experimental and theoretical studies to date both provide strong evidence for a preprotonation mechanism [i.e., the enzyme-bound substrate exists in the protonated form (SH^+)].[12–19] In addition, these novel substrates have an advantage over folate in computational studies in that they do not have the bulky *p*-methylaminobenzoyl-L-glutamate side chain and are therefore more amenable to computation using quantum mechanical (QM) methods. Consequently taking advantage of this suitability for computational studies, in the present work we used combined QM and molecular mechanics (QM/MM) methods to compute the free energy change for the hydride-ion transfer step between NADPH and one such substrate, 8-methylpterin.

The QM/MM method can be used to study chemical processes in condensed phases, such as substrate or inhibitor binding to enzymes and biological catalysis. However, due to the relatively high cost of performing QM calculations within molecular dynamics (MD) simulations, only a relatively small part of the system can be treated quantum mechanically; the remaining larger part of the system is treated using MM potentials. This requirement in itself complicates the implementation of the QM/MM methods because, if the boundary between the QM and MM regions crosses covalent bonds, the restriction of the quantum region to a small number of atoms may affect the chemical accuracy. Note that if covalent bonds cross the boundary between the QM and MM regions, so-called ''link'' atoms[20] must be introduced to satisfy valence requirements and effec-

tively model the actual system. Thus, a realistic model for the reacting system must be developed. The development of QM/MM potentials for describing the interacting systems was partly addressed in previous studies.[20–23]

Another important consideration is the method used to compute properties of the reacting system. With MD simulation methods it is possible, at least in principle, to compute Boltzmann-averaged quantities for direct comparison with experiments. However, for very large and dynamically complex systems, the errors in absolute energies arising from incomplete sampling of configuration space can make the calculation of such small energy differences unreliable. Fortunately, it is possible to calculate free energy differences directly using free energy perturbation (FEP) methods[24] in which energy differences between similar states are computed at defined points along the reaction coordinate. In the present study we applied the FEP method to calculate the free energy change for the hydride-ion step.

The total free energy differences between reactant, product, and transition states does not provide molecular-level information on the catalytic mechanism. Consequently, to understand the role of the enzyme and the relationships between protein structure and function in catalysis, obviously it would be desirable to decompose the free energies into different components. It should be noted, however, that the values obtained for component free energies will, in general, depend on the path taken.[25–28] The exact minimum energy path between reactant and product passes through a first-order saddle point (i.e., the transition state structure) on the free energy surface. However, because the free energy is a state function, we are not restricted to the minimum energy path but are free to explore other pathways in order to determine the free energy of reaction. Thus, the path dependence of the free energy components could be tested by decomposition over different reaction pathways. Also, the free energies obtained from multiple pathways, including the reverse pathways, give an indication of the quality of configuration space sampling and a test of convergence of the FEP calculations.

Methods

COMBINED QM / MM METHODS

The total potential energy (E) of a system partitioned into QM and MM groups of atoms may be written in the general form[5]

$$E = E_{QM} + E_{MM} + E_{QN/MM}, \quad (1)$$

where E_{QM} is the Hartree–Fock self-consistent field (SCF) energy of the quantum system, E_{MM} is the energy of the MM part of the system, and $E_{QM/MM}$ is the interaction energy between the QM and MM parts of the system. When performing MD simulations using QM/MM potentials, we are generally restricted to semiempirical QM methods for reasons of computational efficiency. In the present study we chose the AM1 semiempirical method.[29,30] The QM/MM interaction energy is written in the usual way as the sum of polar (electrostatic, ele) and nonpolar (van der Waals, vdW) terms,

$$E_{QM/MM} = E_{ele} + E_{vdW}, \quad (2)$$

where E_{ele} is the semiempirical approximation for the interaction between the nuclei and electrons of the QM system and the atomic charges of the MM system. The expressions for E_{ele} were given previously and include the effects of polarization of the QM region by the MM region.[23,31] Note, however, that we do not include terms that explicitly describe the polarization of the MM region. The E_{vdW} term is expressed simply as the sum over pair-potential functions from the MM part of the force field.

The system is initially built in the standard way for an MM/MD calculation from a number of relatively small fragments or ''residues.''[32] The 8-methylpterin molecule is residue 1, residue 2 is a CH_3CO C-terminal capping group, residues 3–189 are the individual amino acids of DHFR, residue 190 is an NH_2 N terminus, residues 191–194 (respectively, D, C, B, and A in Fig. 1) constitute NADPH, and residues 195 and above are water molecules. The computations were performed using Molecular Orbital Programs for Simulations (MOPS).[33] All residues are assumed to be MM unless specified as QM (i.e., the QM region in MOPS can only be defined on a residue by residue basis). In the present study, the 8-methylpterin molecule and the four residues of the NADPH molecule, as shown in Figure 1, were included in the QM region in the largest QM-region model. Several other QM/MM models for the cofactor were also defined by restricting the QM region to contain only residues A; A and B (AB); or A, B, and C (ABC). The cofactor residues that do not

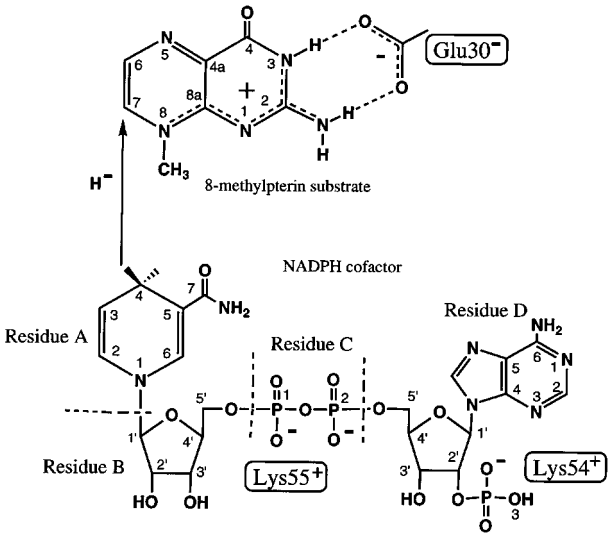

FIGURE 1. 8-Methylpterin substrate and NADPH cofactor indicating the main interactions with charged side chains. The cofactor is partitioned into residues A (nicotinamide), B (ribose), C (phosphate), and D (adenine) for the purposes of defining the QM / MM boundary. The hydride-ion transfer takes place between C4 of the nicotinamide ring of NADPH and C7 of the substrate.

form part of the QM region, all of the protein residues, and all water molecules form the MM region. Where a covalent bond in the NADPH cofactor molecule crosses the boundary between the QM and MM regions, link atoms F, CH_3, or H were used. Thus, we denote the possible QM parts of the cofactor as A-L, AB-L, ABC-L, or ABCD where the link atom L may be F, CH_3, or H. Note that link atoms do not interact directly with the MM atoms through explicit terms in the Hamiltonian, but they are required only to satisfy valency requirements of the QM fragment. Terms in the MM force field describing bonds, angles, and dihedrals that include both QM and MM atoms are retained; the nonbonded electrostatic and vdW terms are treated in the usual way as described in eq. (2). In the present MD simulations the link atoms are treated in the same way as all other atoms: their dynamical behavior is determined by the forces acting on them according to classical mechanics. Although this approach introduces artificial degrees of freedom, it is easy to implement and provides a consistent treatment of the total energy and gradient that is required for solving the classical equations of motion in MD. To minimize the risk of any adverse affects on the calculation of reaction free energies, it is important that the link atoms are not located too close to the reaction center.

The crystal structure of the DHFR.$NADP^+$.biopterin complex[34] reveals a strong H-bond interaction between the 2′-phosphate group and the neighboring 3′-OH group with an O—O separation of 2.6 Å. During the MD simulations in which all cofactor residues are treated by QM, the H-bonded oxygen of the 2′-phosphate group abstracts a proton from the 3′-OH group (residue D), indicating little or no activation barrier for the proton transfer at the AM1 level. However, it was found that this proton transfer could be prevented by a single protonation at the H-bonded oxygen of the 2′-phosphate group (Fig. 1). Consequently, although there is experimental evidence for a dianion 2′-phosphate in the cofactor bound to DHFR,[35] the 2′-phosphate group was treated as a monanion in all simulations. Note that we have also used such a protonation in a previous MD/FEP study of NADPH binding to DHFR.[36]

The force field of Weiner et al.[37,38] (i.e., the AMBER force field) was used for the protein, while the TIP3P model parameters were used for water.[39] For protein residues within 8 Å of the substrate molecule, the ''all atom'' force field[38] was used; for the remainder of the protein residues we used the ''united atom'' force field[37] in which hydrogens attached to carbon are not explicitly defined. The vdW parameters for the substrate have also been selected from the force field of Weiner et al.[38] The NADPH parameters for the MM part of the force field are described elsewhere.[36] For the QM systems all vdW parameters are the same as those used in the MM force field except that the radii of atom types CA are increased to 2.05 Å.[23] Note that the 10−12 potential parameters for the QM/MM interactions given in Table I are not the same as in the AMBER force field; they were chosen from an optimized set obtained from a previous QM/MM study of solute−solvent interactions.[23]

FEP METHODS

The free energy differences between reactant and product states were calculated using the FEP approach.[24] Thus, we are interested in the energy as a function of some predefined reaction coordinate vector **r**. The Helmholtz free energy change ΔA from **r** to $\mathbf{r} \pm \Delta\mathbf{r}$ is then given by the perturba-

TABLE I.
Parameters R_e for 10 – 12 H-Bond Pair Potentials in AM1 QM / MM Model.

QM Residue	QM Group	H Bond (QM ⋯ MM)	R_e (Å)[a]
Substrate	>C=O	O ⋯ H	2.55
	—NH_2	H ⋯ O	2.15
	>N—H	H ⋯ O	2.15
A (nicotinamide)	>C=O	O ⋯ H	2.70
	—NH_2	H ⋯ O	2.10
B (ribose)	—O—H	O ⋯ H	2.35
		H ⋯ O	2.10
C (phosphate)	P—O	O ⋯ H	2.65

[a] See ref. 23 for the definition of R_e. The coefficients of the 10 –12 terms are given by $C_{12} = 0.05R_e^{12}$ and $C_{10} = 0.06R_e^{10}$.

tion formula

$$\Delta A(\mathbf{r}, \mathbf{r} \pm \Delta\mathbf{r}) = -(RT)^{-1} \times \ln\langle \exp[-(E(\mathbf{r} \pm \Delta\mathbf{r}) - E(\mathbf{r}))/RT]\rangle_{\mathbf{r}} \quad (3)$$

where the ensemble average is over the state defined at some point $\mathbf{r}$ on the reaction coordinate. We estimated the free energy change for each window by taking the mean of plus and minus perturbations:

$$\Delta A_i = \{\Delta A(\mathbf{r}_i, \mathbf{r}_i + \Delta\mathbf{r}) - \Delta A(\mathbf{r}_i, \mathbf{r}_i - \Delta\mathbf{r})\}/2. \quad (4)$$

The cumulative total (i.e., the sum over all windows i) gives the free energy of reaction. The above FEP formula [eq. (3)] may also be written for the individual energy components E_{QM} and $E_{QM/MM}$ in eq. (1),

$$\Delta A_{QM}(\mathbf{r}, \mathbf{r} \pm \Delta\mathbf{r}) = -(RT)^{-1} \ln\left\langle \exp\left[-\left(E_{QM}(\mathbf{r} + \Delta\mathbf{r}) - E_{QM}(\mathbf{r})\right)/RT\right]\right\rangle_{\mathbf{r}} \quad (5a)$$

$$\Delta A_{QM/MM}(\mathbf{r}, \mathbf{r} \pm \Delta\mathbf{r}) = -(RT)^{-1} \ln\left\langle \exp\left[-\left(E_{QM/MM}(\mathbf{r} \pm \Delta\mathbf{r}) - E_{QM/MM}(\mathbf{r})\right)/RT\right]\right\rangle_{\mathbf{r}}. \quad (5b)$$

The total free energy change may be expressed as the sum of component free energies,

$$\Delta A = \Delta A_{QM} + \Delta A_{QM/MM}. \quad (6)$$

Further, the QM/MM component in eq. (6) can be written analogously as the sum of electrostatic and vdW components as defined in eq. (2),

$$\Delta A_{QM/MM} = \Delta A_{ele} + \Delta A_{vdW}. \quad (7)$$

A formal correction to the free energy ΔA_{QM} would be expected to arise from the distance constraints[40] that are imposed in order to define points along the reaction coordinate. This correction amounts to the difference in the work required to maintain the constraints in the reactant and product states. In the present study, however, similar H—C bonds are being constrained in both reactant and product states. Moreover, FEP calculations by Ho et al.[41] yield good agreement with experimental free energies for hydride-ion transfer in solution without invoking such corrections. Consequently, we also did not attempt to calculate small constraint-dependent free energy corrections.

MD SIMULATIONS AND FEP CALCULATIONS

The starting coordinates for the present simulations were taken from a previous study of solvated DHFR ternary complexes.[42] Briefly, the X-ray crystal structure of the avian DHFR.$NADP^+$.biopterin complex[34] provided the starting coordinates for the study.[42] A 16-Å radius active-site dynamics zone was then solvated with a 22-Å shell containing 392 dynamical water molecules. Details of the initial setup of the system and solvation treatment are given elsewhere.[42] Starting from a coordinate set obtained from the previous MD study of the DHFR.NADPH.8-methylpterin complex, a C7 (substrate) and C4 (cofactor) separation of 3.5 Å was imposed using the SHAKE algorithm.[43] Using the MM-only force field, the complex was initially energy minimized for 100 cycles of conjugate gradients, followed by a 150-ps equilibration period of MD simulation. We used the constant temperature algorithm of Berendsen et al.[44] to perform the MD simulations with a time step of 0.001 ps and an H mass of 3 amu. The temperature was set at 300 K with a relaxation time of 0.1 ps. A 9-Å cutoff was used for the neglect of interactions within the MM region, and there were updates every 50 time steps for the interaction list.

Simulation conditions for the QM/MM systems were set to the same as those used above, except that no cutoff was applied to the nonbonded interactions between the QM and MM regions [eq. (2)]. The two interatomic r(C4-H) and r(C7-H) dis-

tances were chosen for the reaction coordinate. The reaction coordinate was maintained at fixed values by application of the SHAKE algorithm.[43] The MD simulations using the QM/MM force field were first carried out for the system in which A-F (i.e., link F) was the QM species. The equilibrated A-F structure was then used as the starting point for the remaining A-L, AB-L, ABC-L, and ABCD QM species. The necessary equilibrations were carried out on all systems before FEP calculations.

For the perturbation equations to be valid, the change in free energy for each $\Delta \mathbf{r}$ should be less than $3RT$. To stay within this limit, the reaction coordinates were divided into windows of width $\Delta \mathbf{r}$ = 0.01 or 0.02 Å, requiring on the order of 160 windows to obtain the total free energy difference between the reactant and product states. With each window consisting of 800 steps of equilibration followed by 800 steps of data collection for calculation of the free energies, the overall simulation time was 256 ps. While some windows were subject to longer simulation times as a test of sampling, we chose a standard of 1.6 ps per window as the best compromise between reliability and practicality. This gave a total sampling time close to the minimum used by Ho et al.[41] in FEP calculations of hydride-transfer reactions in solution. As a further test of sampling, some FEP calculations were performed following the reverse pathway (i.e., going from product to reactant). To treat the entire cofactor (residues ABCD) by QM and complete the FEP calculation within a reasonable time frame, we exploited the availability of machines with multiple processors. We performed a number of FEP calculations concurrently on separate processors by dividing the reaction path into segments of 20 windows each (i.e., allowing for a maximum of eight processors). The total free energy change is then the sum over free energies obtained from the individual FEP calculations. To ensure that the structures at the beginning of each segment were properly equilibrated, a minimum of 20 ps of MD was allowed before the FEP calculations were started.

Results and Discussion

The free energy differences between reactant and product states were computed along the paths shown in Figure 2. The most direct path between the reactant and product states is denoted β, whereas path β' begins and ends the same as path

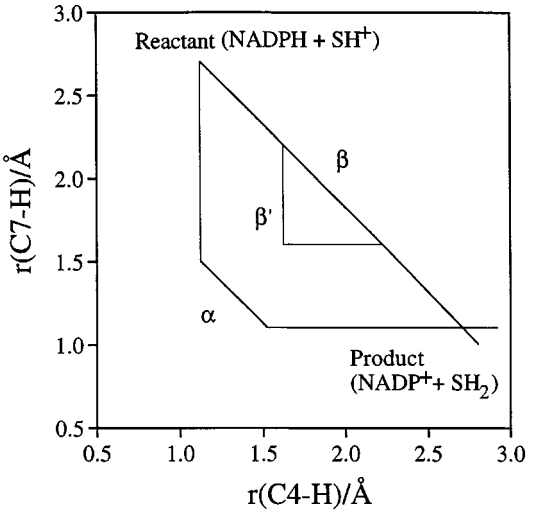

FIGURE 2. Reaction pathways used in the FEP calculations. The pathways are defined by allowing increments (windows) Δr of absolute value 0.00, 0.01, or 0.02 Å such that $|\Delta r(\text{C4-H})| + |\Delta r(\text{C7-H})| = 0.02$ Å. The reactant state is defined at r(C4-H) = 1.125 Å and r(C7-H) = 2.705 Å, and the product state is defined at r(C4-H) = 2.704 Å and r(C7-H) = 1.105 Å.

β but avoids the high energy barrier associated with β. Based on previous AM1 studies[45] of gas-phase analogue reactions, the r(C4-H) and r(C7-H) bond lengths at the TS are expected to be between 1.3 and 1.5 Å. Thus, path α closely follows the minimum free energy path and consequently approaches the TS. Therefore, a good estimate of the free energy of activation ($\Delta A^{\ddagger}$) can be obtained from following the FEP calculations along path α. In order to calculate a free energy of reaction and $\Delta A^{\ddagger}$, it was necessary to define reactant and product states in terms of specific values of r(C4-H) and r(C7-H). For the reactant state based on the simulation using A-F as the QM species, a local minimum was determined at r(C4-H) = 1.125 Å and r(C7-H) = 2.705 Å. Similarly the product state was characterized by r(C4-H) = 2.704 Å and r(C7-H) = 1.105 Å. These values were not significantly different for the other A-L species.

The free energies for the hydride-ion transfer from the cofactor to the substrate are given in Table II. The activation free energies estimated from path α are given in parentheses. The component (QM and QM/MM) free energies calculated using eq. (5) add to give a total free energy that is very close to the value calculated using eq. (3). The QM/MM components for a given reaction show only small variations, depending on the reaction pathway, which is consistent with sampling errors

TABLE II.
Free Energy Changes (kcal / mol) for Hydride-Ion Transfer Pathways α, β, and β′ Described in Figure 2.

QM Species[a]	Path[b]	ΔA (kcal / mol) QM	QM / MM	Electrostatic[c]	Total[d]	Total[e]
A-F	α	21.66	−4.36	−4.29	17.30	17.30 (39.2)
	β	18.26	−5.36	−5.22	12.90	12.92
A-CH_3	α	8.91	−7.41	−7.33	1.50	1.50 (34.0)
	β	4.98	−6.89	−6.93	−1.91	−1.87
A-H	α	11.42	−4.67	−4.42	6.75	6.76 (30.6)
	β	9.78	−5.67	−5.60	4.11	4.12
	β′	9.47	−3.50	−3.34	5.97	5.99
AB-H	α	−20.06	0.71	0.81	−19.35	−19.35 (21.9)
	β′	−13.74	−1.32	−1.25	−15.06	−15.04
ABC-H	α	5.26	9.24	9.52	14.50	14.54 (33.0)
	α (reverse)	−9.36	−6.64	−6.87	−16.01	−16.00
	β′	15.64	9.86	10.07	25.50	25.56
ABCD	α	−0.15	9.31	9.56	9.16	9.20 (29.3)
	α (reverse)	0.77	−8.28	−8.57	−7.51	−7.53

[a] QM parts of the cofactor A-L, AB-L, ABC-L, or ABCD, where the link atom L may be F, CH_3, or H. Note the entire substrate molecule is treated by QM.
[b] The reverse is the free energy obtained by following the path from products to reactants.
[c] The electrostatic / polarization component of the free energy as defined in eq. (7).
[d] The total obtained by summing the QM and QM / MM components.
[e] The total calculated from eq. (3). The results in parentheses for path α are the estimates of the activation free energy.

rather than indicative of component free energies not behaving as state functions. The corresponding variations in the QM components are generally much larger, but they are also consistent with sampling errors when it is considered that the dynamical fluctuations in the QM energy terms are much greater than for the QM/MM terms. The difference in total free energy between the forward and reverse pathways is about 2 kcal/mol, whereas differences in the corresponding component free energies may be as high as 4 kcal/mol (QM component for ABC-H).

As may be expected for a redox reaction, the change in the QM/MM component is dominated by the electrostatic as opposed to vdW type of interactions. The QM/MM components range from ca −4 to −7 kcal/mol for the A-L species, favoring the formation of the positive charge on the nicotinamide residue. However, as more of the NADPH cofactor residues are treated by QM, the QM/MM component of the free energy becomes more positive (i.e., the enzyme charge distribution does not favor the shift of net positive charge from the substrate to the cofactor). In simulations with all NADPH residues treated by QM, the interactions with the enzyme plus solvent (the QM/MM terms) contribute 10 kcal/mol to the free energy change. The free energy changes with respect to reaction coordinate α are shown in Figure 3 for A-H and ABCD reactions. Although it is clear from the ABCD reaction (Fig. 3) that the effect of the solvated-protein environment (i.e., the QM/MM contribution) is to increase the free energy at the TS, the total activation barrier of ca. 30 kcal/mol is similar to that calculated for analogue gas-phase reactions.[45] We note, however, that quantum dynamical effects are likely to reduce this estimate of $\Delta A^{\ddagger}$.[46, 47]

Note that not all of the reactions involving cofactor fragments terminated with link atoms provide an accurate model of the reaction compared with that where the entire NADPH cofactor is treated quantum mechanically. There are often large differences in the QM component free energies that may be rationalized in terms of the electronic structures of the fragments. The charges obtained by summing the atomic partial charges on the QM residues are given in Table III. Also included are the corresponding residue charges obtained for the unpolarized state (i.e., the isolated substrate and cofactor). These charges are obtained simply by recalculating atomic charges from an SCF calculation in which the QM/MM terms are neglected, thus giving an indication of the effect the MM atoms have on the QM residue charges. Because both F and CH_3 are strongly electron

withdrawing in the field of the MM atoms, the nicotinamide residue (A) is left with a net positive partial charge in A-F and A-CH_3 whereas a net negative partial charge is obtained in A-H and when all the NADPH residues (i.e., ABCD) are treated by QM. On oxidation of the cofactor (product state) the positive charge on residue A in A-L is overestimated compared with residue A in ABCD, but this overestimation is least when the link atom is H. Thus, an electron donating group (e.g., H) would be the preferred choice for the link in A-L. In contrast, residue B (ribose) in AB-H has a partial charge opposite in sign to the one in ABCD. In the unpolarized state the H link atom in AB-H is electron donating. However, on polarization, residue B is significantly depleted of electrons to such an extent that, in the product state, most of the positive charge resides on residue B rather than A. The situation is somewhat improved in ABC-H, where the partial charges on residues A and B are similar to those found on the corresponding residues in ABCD. Although the charge on C is grossly overestimated in ABC-H, its value does not significantly change on oxidation of the cofactor. Note that the QM species with charge distributions similar to corresponding residues in ABCD also tend to yield similar free energies (Table II).

While analysis of the QM fragment charge distributions may be useful for determining the appropriate QM species and link atoms as a model for a particular reaction, it should be remembered that different polar environments may significantly alter the distribution of electronic charge. It is known from other QM/MM studies[48] that charges on MM atoms close to the QM/MM boundary tend to induce large polarization between QM residues and the link atoms. In the present study also, electronic polarization appears much larger where link atoms are used. This is particularly noticeable where the QM/MM boundary divides highly polar groups. However, the link

TABLE III.
Net Charges on QM Residues in Reactant and Product States.

QM Species[a]	Residue	Reactant[b]	Product[b]
A-F	Substrate	1.00 (1.00)	−0.09 (−0.03)
	A (nicotinamide)	0.25 (0.06)	1.32 (0.97)
	F (link)	−0.25 (−0.06)	−0.23 (0.06)
A-CH_3	Substrate	1.00 (1.00)	−0.10 (−0.02)
	A (nicotinamide)	0.21 (−0.02)	1.36 (0.85)
	CH_3 (link)	−0.21 (0.02)	−0.26 (0.17)
A-H	Substrate	1.00 (1.00)	−0.06 (−0.02)
	A (nicotinamide)	−0.09 (−0.21)	0.98 (0.73)
	H (link)	0.09 (0.21)	0.08 (0.29)
AB-H	Substrate	1.00 (1.00)	−0.10 (−0.06)
	A (nicotinamide)	−0.20 (−0.16)	0.10 (0.06)
	B (ribose)	0.33 (−0.04)	1.32 (0.77)
	H (link)	−0.13 (0.21)	−0.32 (0.23)
ABC-H	Substrate	1.00 (0.21)	−0.04 (−0.09)
	A (nicotinamide)	−0.12 (−0.15)	0.77 (0.11)
	B (ribose)	−0.30 (−0.48)	−0.14 (−0.41)
	C (phosphate)	−2.43 (−0.94)	−2.47 (−0.99)
	H (link)	0.84 (0.35)	0.88 (0.38)
NADPH	Substrate	1.00 (0.96)	−0.05 (−0.07)
	A (nicotinamide)	−0.13 (−0.19)	0.74 (0.71)
	B (ribose)	−0.47 (−0.48)	−0.31 (−0.36)
	C (phosphate)	−0.74 (−0.63)	−0.74 (−0.64)
	D (adenine)	−1.65 (−1.65)	−1.63 (−1.65)

Values are in atomic units.
[a]QM parts of the cofactor A-L, AB-L, ABC-L, or ABCD where the link atom L may be F, CH_3, or H. Note the entire substrate molecule is treated by QM.
[b]Charges obtained over 5 ps of coordinate averaging. Results in parentheses are for the unpolarized state (see text).

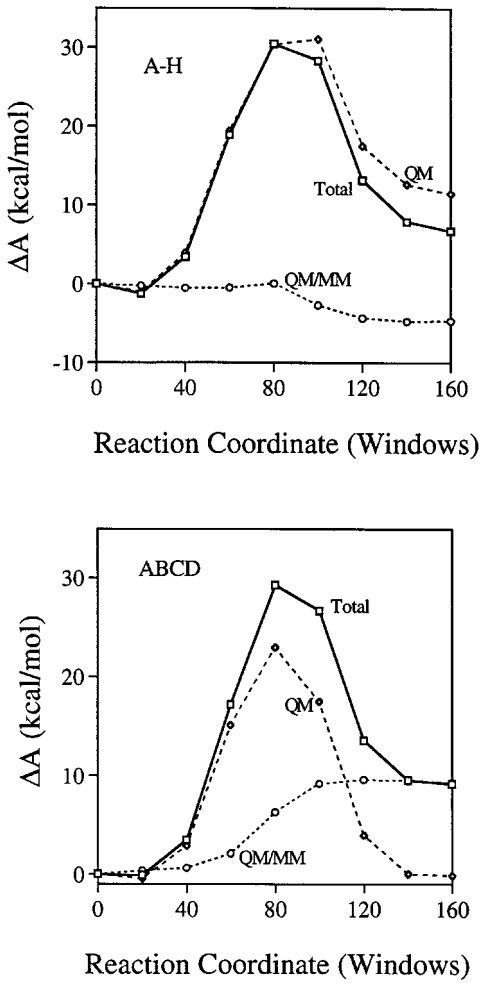

FIGURE 3. Free energy changes (kcal / mol) as a function of the reaction coordinate (number of windows in the FEP calculation) along pathway α for QM species A-H and ABCD. The total free energy is calculated using eq. (3), while those for the QM and QM / MM component free energies are calculated using eq. (5).

atom charges are similar in reactant and product states of A-L and ABC-H; consequently, the polarization would not be expected to have a major effect on the free energy. As a simple test of the effect this polarization has on the free energy changes, we computed the free energy corresponding to the unpolarized QM fragments using the statistics obtained in the usual polarized calculations. The difference in the free energies was found to be not more than 1 kcal/mol.

Although residue C forms part of the 5′-phosphate group and has a formal charge of -2, the calculated value is only ca. -0.7, the excess negative charge being distributed over residues B and D. The charge on the adjoining oxygen O5′ atoms of residues B and D is also ca. -0.7, so that the -2 formal charge appears largely delocalized over the 5′-phosphate (P_2O_7) group. On polarization of the cofactor, ca. 0.1 of an electron moves from the nicotinamide end of the cofactor to residue C. Bajorath et al.[10] also studied the charge distribution in NADPH bound to *Escherichia coli* DHFR using a different computational model and found a shift of 0.7 electrons. Whether this difference in polarization was due to species differences in DHFR or methodological differences merits further study.

The H-bond interactions between QM and MM regions in the reactant and product states are listed in Table IV. The QM/MM model is capable of reproducing the H bonding observed in the X-ray crystal structure. In both reactant and product states the amide side chain of the nicotinamide ring is H bonded to the backbone of Ala19 and Ile16, and the side chain is rotated about the C5–C7 bond (Fig. 1) ca. 30° from planar. Although the AM1 model underestimates the increase in energy for out of plane rotation of the side chain, we would not expect such a rotation to be prohibited at a higher level of QM theory.[49] While the H bonds between nicotinamide and DHFR fix the ring in the general position that is necessary for hydride-ion transfer, rotation about the C5–C7 bond may allow some additional flexibility for improving the relative orientation of the substrate and nicotinamide. This configuration is very different from gas-phase analogue reactions in which the side chain may be involved in H-bonded complex formation between reacting fragments.[45] Thus, the significance of the nicotinamide side-chain orientation in catalysis seems to be mechanical rather than electronic. In the substrate-binding site, very strong H bonding with Glu30 is observed in both the reactant and product states. Other H bonds are of lesser importance for catalysis but contribute to the overall binding of the cofactor. The polar groups of cofactor residues C and D and the positively charged side chains of neighboring Lys54 and Lys55 residues are extensively solvated by surface water. Hence, the MD trajectories reveal a significant number of labile H bonds in the cofactor binding pocket. In the reactant stage, the H-bond length with Lys55 exhibits a large standard deviation and the H bond is readily exchanged for

H bonds with water molecules, while there is an absence of H bonding with Lys54 in the products. This dynamical behavior is consistent with general NMR studies of protein–solvent interactions that have determined subnanosecond residence times for water molecules on protein surfaces.[50,51]

To understand the origin of the 10 kcal/mol free energy change we calculated the mean difference in the QM/MM interaction term between reactant and product states as a function of residue number. The results are given in Figure 4 for QM species A-H and ABCD. As indicated in the figure, there are a number of interactions between the QM residues and charged MM residues that contribute significantly to the energy difference. Of those energetically significant interactions, only Glu30, Lys54, and Lys55 are associated with H bonding between QM and MM atoms. Obviously, A-H gives rise to contributions from the interaction of QM residue A with MM cofactor residues B, C, and D (residue numbers 193–191). In ABCD these contributions are accounted for in the purely QM energy term. The contribution from Lys18 is largely cancelled by the negative contributions from Gly17 and Asp19. In fact, the overall effect of the remaining non-H-bond contributions, which are long-range electrostatic in character, appears to be minimal due to cancellation. The much larger contribution from Lys54 for the interaction with ABCD is clearly due to the loss of H bonding with the 2′-phosphate group of QM residue D that is observed in the product state. The relatively large energy difference for the interaction between ABCD and Lys54 (Fig. 4) is due to H-bonding differences between reactant and product states (Table IV). Note, however, that these differences do not arise in the MD/FEP calculations because the changes in energies used to compute free energies are always obtained between similar states [eqs. (3), (5)], i.e., for any step generated in the MD simulation Lys54 is either H bonded or not H bonded to the cofactor. The only significant term remaining is the interaction of the QM residues with Glu30. Because the magnitude of this contribution is relatively large (35–40 kcal/mol) and not dependent on the QM model used for the cofactor

TABLE IV.

Hydrogen Bond Lengths (Å) in Reactant and Product States for Interactions between QM and MM Atoms for Simulation in which All Cofactor Residues (ABCD) Are Treated by QM.

QM Residue	QM atom[a]	MM Atom	Reactant[b]	Product[b]
Substrate	H3	Glu30 OE2	1.54 (1)	1.59 (2)
	H2a	Glu30 OE1	1.54 (1)	1.58 (1)
	H2b	OW	1.80 (6)	2.98 (88)
	O4	HW	1.96 (7)	1.98 (14)
A (nicotinamide)	O7	Ala9 HN	2.02 (2)	2.00 (5)
	H7a	Ala9 O	1.81 (4)	1.74 (5)
	H7b	Ile16 O	1.84 (3)	1.92 (13)
B (ribose)	HO2′	Lys18 O	1.72 (4)	1.69 (4)
	O2′	HW	1.81 (10)	1.88 (12)
	HO3′	Lys18 O	1.95 (5)	2.11 (40)
	O3′	Ser59 HG	2.49 (47)	2.92 (46)
C (phosphate)	O1a	Lys55 HNZ	2.45 (73)	2.00 (6)
	O1a	HW	1.95 (8)	2.09 (15)
	O1b	Thr118 HN	2.09 (4)	2.18 (14)
	O1b	HW	2.00 (20)	2.02 (13)
	O2a	Thr56 HN	1.91 (1)	1.93 (6)
	O2b	HW	1.96 (13)	1.92 (11)
D (adenine)	Oa	Lys54 HNZ	1.90 (4)	4.89 (86)
	Oa	HW	1.92 (3)	1.93 (10)
	Ob	Arg77 HN	2.03 (7)	2.05 (4)
	O3	HW	2.28 (59)	1.95 (15)
	O3′	HW	1.92 (14)	—
	HN6	OW	1.95 (12)	1.97 (12)

[a] See Figure 1 for atom labels.

[b] The H-bond lengths were obtained over 5 ps of coordinate averaging. Standard deviations are given in parentheses.

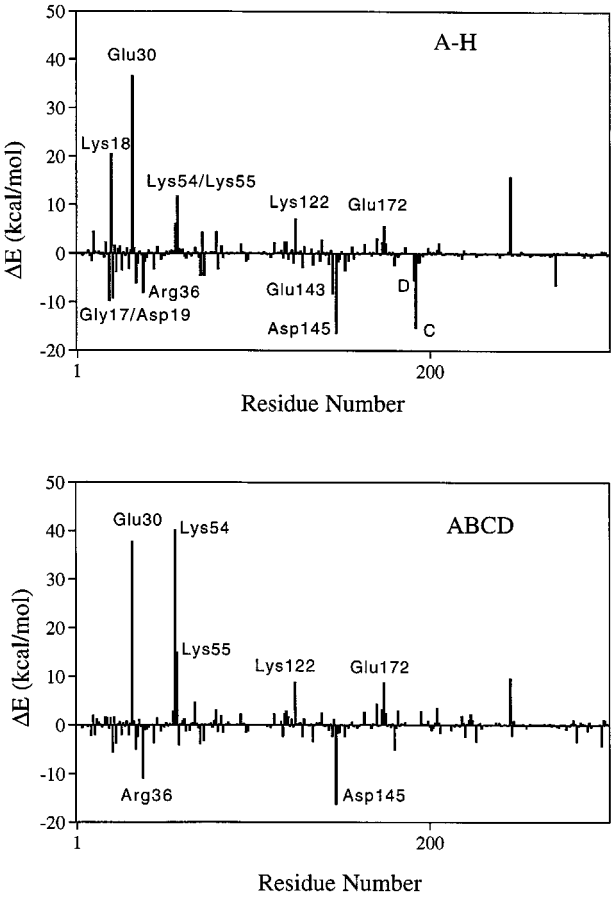

FIGURE 4. Differences in the QM / MM energy between reactant and product states as a function of residue number and based on a 5-ps interval of coordinate averaging. The 8-methylpterin molecule is defined as residue 1, residue 2 is a CH_3CO C-terminus group, residues 3 –189 are the amino acids, residue 190 is an NH_2 N terminus, residues 191 –194 (respectively, D, C, B, and A in Fig. 1) constitute NADPH, and residues 195 and above are water molecules.

(i.e., A-H or ABCD), it can be attributed to the strong interaction of Glu30 with the substrate. The increase in the QM/MM component of the free energy may thus be rationalized largely in terms of the decrease of electrostatic binding energy in the H-bond interaction (Fig. 1) between the charged Glu30 side chain and the charged substrate when going to uncharged product.

Conclusions

In the present study of the enzyme-catalyzed hydride-ion transfer step between NADPH cofactor and 8-methylpterin substrate, we partitioned the NADPH cofactor into a number of subunits (residues) to be treated by either QM or MM force fields. With the appropriate choices of QM cofactor residues and link atoms, we demonstrated the feasability of modeling the full QM treatment of the NADPH cofactor in terms of both the electronic distributions of the QM residues and the free energies calculated using the FEP method. Clearly, it is preferable that the QM/MM boundary be located far from the reaction center. However, if this is not practicable, the choice of link atoms becomes an important consideration. For the smallest QM fragment (nicotinamide) used in the present study, we found that an H atom was the most appropriate choice for the link due to its electron donating properties.

Assuming a preprotonation mechanism (i.e., one in which the active form of the substrate is protonated), we calculate a barrier of 30 kcal/mol for the hydride-ion transfer step. However, we regard this value as an upper bound estimate because quantum dynamical effects, which have been neglected in the present study, are likely to reduce the activation free energy.[46,47] By performing the FEP calculations over different pathways, we verified that the total free energy changes can be discussed in terms of QM and QM/MM component free energies. The total free energy change of 10 kcal/mol for the transfer can be attributed largely to the QM/MM contribution arising from the specific H-bond interaction of the negatively charged side chain of Glu30 with the protonated substrate molecule. Thus, the hydride ion is transferred at the expense of a reduction in the binding energy between substrate and enzyme.

Acknowledgments

We gratefully acknowledge large grants of computer time from the Australian National University (ANU) Supercomputer Facility on the Fujitsu VPP300 and SGI PowerChallenge systems. The research was funded by an ANU Strategic Development Grant.

References

1. A. Warshel, Computer Modeling of Chemical Reactions in Proteins and Solutions, Wiley, New York, 1991.
2. A. Warshel, *Curr. Opin. Struct. Biol.*, **2**, 230 (1992).
3. J. Aqvist and A. Warshel, *Chem. Rev.*, **93**, 2523 (1993).
4. K. M. Merz, *Curr. Opin. Struct. Biol.*, **3**, 234 (1993).

5. J. Gao, *Rev. Comput. Chem.*, **7**, 119 (1996).
6. W. R. Cannon, B. J. Garrison, and S. J. Benkovic, *J. Am. Chem. Soc.*, **119**, 2386 (1997).
7. J. E. Gready, *Biochemistry*, **24**, 4761 (1985).
8. J. Bajorath, J. Kraut, Z. Li, D. H. Kitson, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA*, **88**, 6423 (1991).
9. J. Bajorath, D. H. Kitson, G. Fitzgerald, J. Andzelm, J. Kraut, and A. T. Hagler, *Proteins: Struct. Funct. Genet.*, **9**, 217 (1991).
10. J. Bajorath, Z. Li, G. Fitzgerald, D. H. Kitson, M. Farnum, R. M. Fine, J. Kraut, and A. T. Hagler, *Proteins: Struct. Funct. Genet.*, **11**, 263 (1991).
11. K. Ramnarayan, F. H. Hausheer, and U. C. Singh, *Chem. Des. Automation News*, September/October, 18 (1993).
12. J. E. Gready, P. L. Cummins, and P. Wormell, *Adv. Exptl. Med. Biol.*, **338**, 487 (1993).
13. V. Thibault, M. J. Koen, and J. E. Gready, *Biochemistry*, **28**, 6042 (1989).
14. P. L. Cummins and J. E. Gready, *J. Comput. Aided Mol. Des.*, **7**, 535 (1993).
15. P. L. Cummins and J. E. Gready, *Proteins: Struct. Funct. Genet.*, **15**, 426 (1993).
16. S. S. Jeong and J. E. Gready, *Eur. J. Biochem.*, **221**, 1055 (1994).
17. M. T. G. Ivery and J. E. Gready, *J. Med. Chem.*, **37**, 4211 (1994).
18. M. T. G. Ivery and J. E. Gready, *Biochemistry*, **34**, 3724 (1995).
19. S. S. Jeong and J. E. Gready, *Biochemistry*, **34**, 3734 (1995).
20. M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.*, **11**, 700 (1990).
21. V. V. Vasilyev, A. A. Bliznyuk, and A. A. Voityuk, *Int. J. Quantum Chem.*, **44**, 897 (1992).
22. P. A. Bash, L. L. Ho, A. D. MacKerell, D. Levine, and P. Hallstrom, *Proc. Natl. Acad. Sci. USA*, **93**, 3698 (1996).
23. P. L. Cummins and J. E. Gready, *J. Comput. Chem.*, **18**, 1496 (1997).
24. M. Mezei and D. F. Beveridge, *Ann. NY Acad. Sci.*, **1**, 482 (1986).
25. P. E. Smith and W. F. van Gunsteren, *J. Phys. Chem.*, **98**, 13735 (1994).
26. S. Boresch, G. Archontis, and M. Karplus, *Proteins: Struct. Funct. Genet.*, **20**, 25 (1994).
27. A. E. Mark and W. F. van Gunsteren, *J. Mol. Biol.*, **240**, 167 (1994).
28. M. Zacharias and T. P. Straatsma, *Mol. Simulation*, **14**, 417 (1995).
29. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
30. M. J. S. Dewar and C. Jie, *J. Mol. Struct.* (*Theochem.*), **187**, 1 (1989).
31. P. L. Cummins and J. E. Gready, *Chem. Phys. Lett.*, **225**, 11 (1994).
32. P. K. Weiner and P. A. Kollman, *J. Comput. Chem.*, **2**, 287 (1981).
33. P. L. Cummins, *Molecular Orbital Programs for Simulations* (*MOPS*), Australian National University, 1996.
34. M. A. McTigue, J. F. Davies, B. T. Kaufman, and J. Kraut, *Biochemistry*, **31**, 7264 (1992).
35. B. Birdsall, G. C. K. Roberts, J. Feeney, and A. S. V. Burgen, *FEBS Lett.*, **80**, 313 (1977).
36. P. L. Cummins, K. Ramnarayan, U. C. Singh, and J. E. Gready, *J. Am. Chem. Soc.*, **113**, 8247 (1991).
37. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
38. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
40. W. F. van Gunsteren, In Computer Simulation of Biomolecular Systems, W. F. van Gunsteren and P. K. Weiner, Eds., ESCOM, Leiden, 1989, p. 27.
41. L. L. Ho, A. D. MacKerell, and P. A. Bash, *J. Phys. Chem.*, **100**, 4466 (1996).
42. P. L. Cummins and J. E. Gready, *J. Comput. Chem.*, **17**, 1598 (1996).
43. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, **34**, 1311 (1977).
44. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
45. P. L. Cummins and J. E. Gready, *J. Comput. Chem.*, **11**, 791 (1990).
46. Y. S. Kong and A. Warshel, *J. Am. Chem. Soc.*, **117**, 6234 (1995).
47. J. K. Hwang, Z. T. Chu, A. Yadav, and A. Warshel, *J. Phys. Chem.*, **95**, 8445 (1991).
48. D. Bakowies and W. Thiel, *J. Phys. Chem.*, **100**, 10580 (1996).
49. P. L. Cummins and J. E. Gready, *J. Mol. Struct.* (*Theochem.*), **183**, 161 (1989).
50. G. Otting, E. Liepinsh, and K. Wuthrich, *Science*, **245**, 974 (1991).
51. J. Qin, G. M. Clore, and A. M. Gronenborn, *Structure*, **2**, 503 (1994).